How-To Library

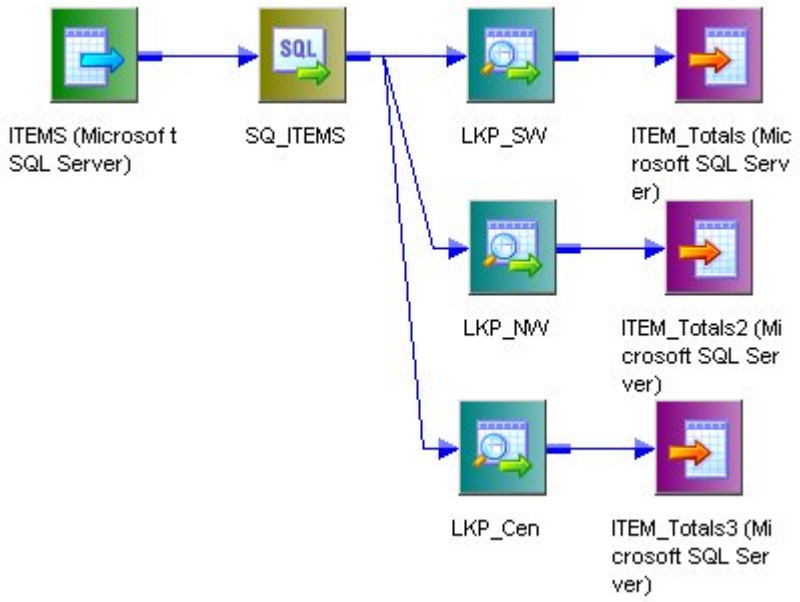
INFORMATICA

Using Persistent Lookup Caches to Increase Lookup Performance

Overview

When you run a session with multiple Lookup transformations, the Integration Service creates caches for the Lookup transformations sequentially. If you have large lookup sources, you can increase performance by creating sessions to build persistent lookup caches concurrently. You can then reuse the persistent caches in the session without the performance impact of building each lookup cache from the database.

The Integration Service builds caches for Lookup transformations sequentially, even when you branch the mapping or add partitions. For example, you have the following Sales mapping:



The structure of the mapping is parallel. However, the Integration Service builds the lookup caches sequentially. Create named persistent caches and use them in the session.

Use persistent lookup caches when you know the lookup table does not change between session runs.

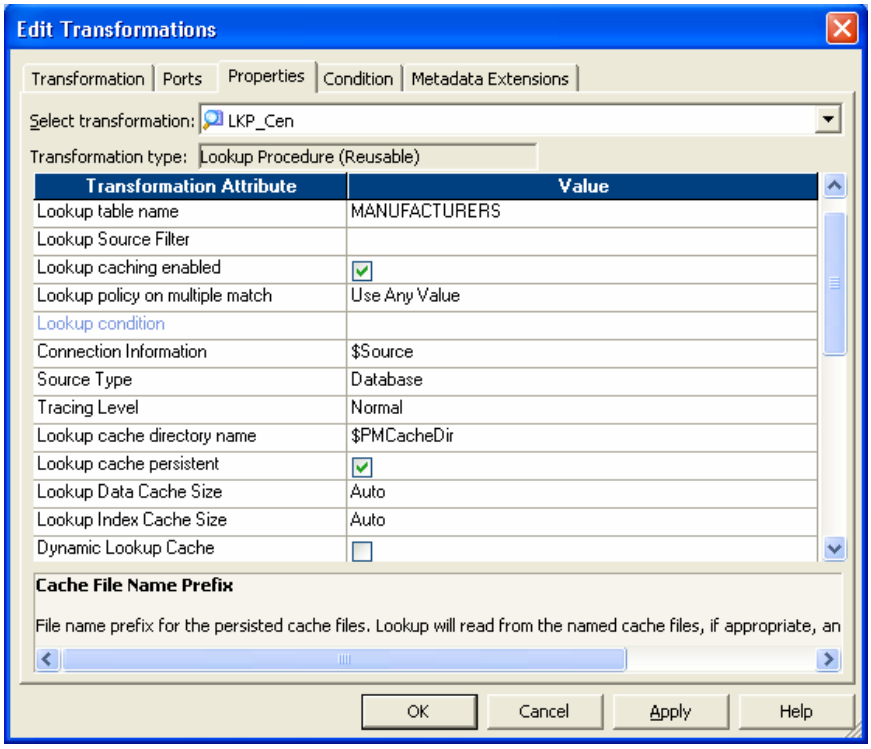
Steps to Create and Reuse Persistent Caches

Complete the following steps to build persistent caches and use them in your session:

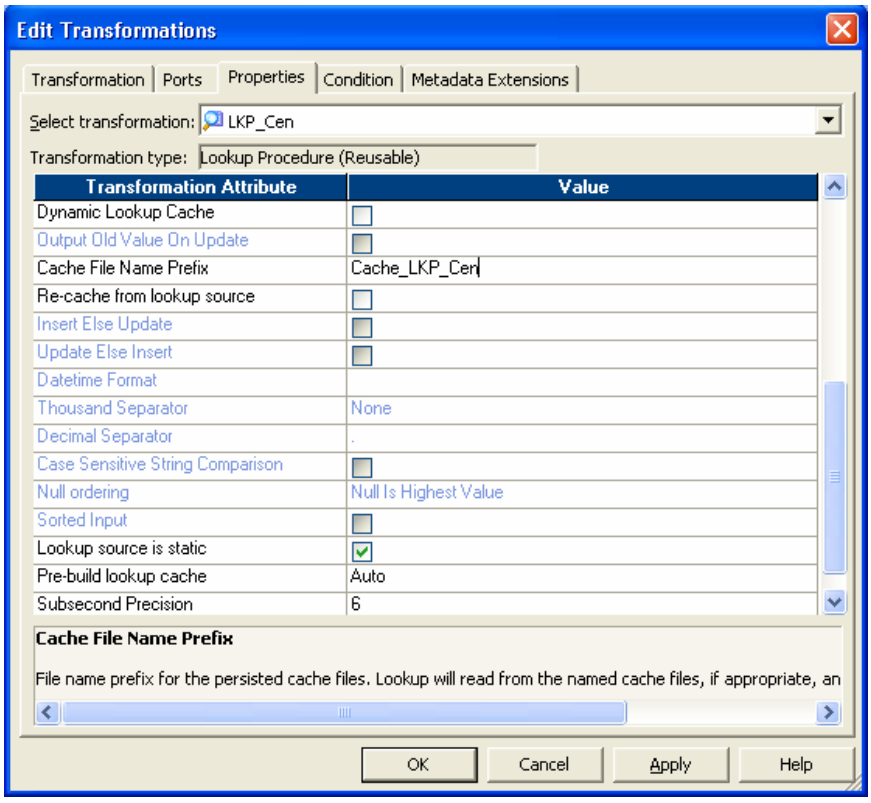
1. Create reusable Lookup transformations. When you create the Lookup transformations, enable persistent caches and specify a cache file name prefix.
2. Create the mappings to build the persistent caches. Create a separate mapping for each Lookup transformation. You use the same reusable Lookup transformations in this mapping that you use in the mapping to process the data.
3. Create the mapping to process the data. Create a mapping that uses all the Lookup transformations you used in step 2. This mapping can use the persistent caches you created in Step 2.
4. Create a session for each of the mappings. Create a separate session for each of the mappings. The Integration Service can process caches in parallel when the Lookup transformations are in different sessions.
5. Create the workflow. Create a workflow to include the sessions that build the persistent caches and the session that processes the data.

Step 1: Create Reusable Lookup Transformations

Create the Lookup transformations you want to use in the mapping. To save and reuse the cache files, enable persistent caches in the Lookup transformation properties.



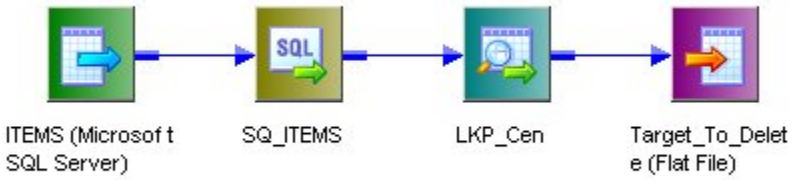
To reuse the cached files across different sessions, specify a cache file name prefix.



Step 2: Create the Mappings to Build the Persistent Caches

To build the persistent caches, create a mapping for each lookup cache and write the data to a flat file target you can discard. When you build the mapping, add a filter to the Source Qualifier so that the Integration Service reads only one

row of data. Later, you create a session based on each of these mappings. For example, you need to create caches for the Sales mapping with the following Lookup transformations: LKP_Cen, LKP_NW, and LKP_SW. Create the following mapping for the LKP_Cen Lookup transformation:

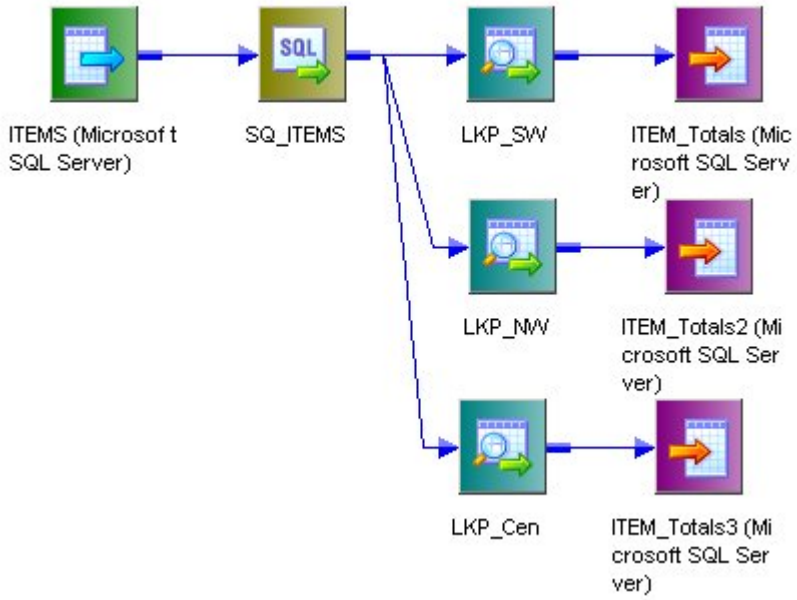


The mapping reads the source, creates the cache for the Lookup transformation and writes to the flat file target.

Create similar mappings for LKP_SW and LKP_NW. Verify that the cache for each Lookup transformation is named and persistent.

Step 3: Create the Mapping to Process the Data

After you create the mapping to build the persistent caches, create the mapping to process the data. Use the Lookup transformations and persistent named caches you created in Step 1. For example, you create the following Sales mapping:



The three Lookup transformations in this mapping are the same ones you used to build the persistent caches.

Step 4: Create a Session for Each of the Mappings

After you create the mappings to build the persistent caches, you need to create a separate session for each of the mappings. For example, you create the following sessions corresponding to each of the mappings you created in Step 2: S_LKP_Cen, S_LKP_NW, S_LKP_SW. These sessions build the persistent caches.

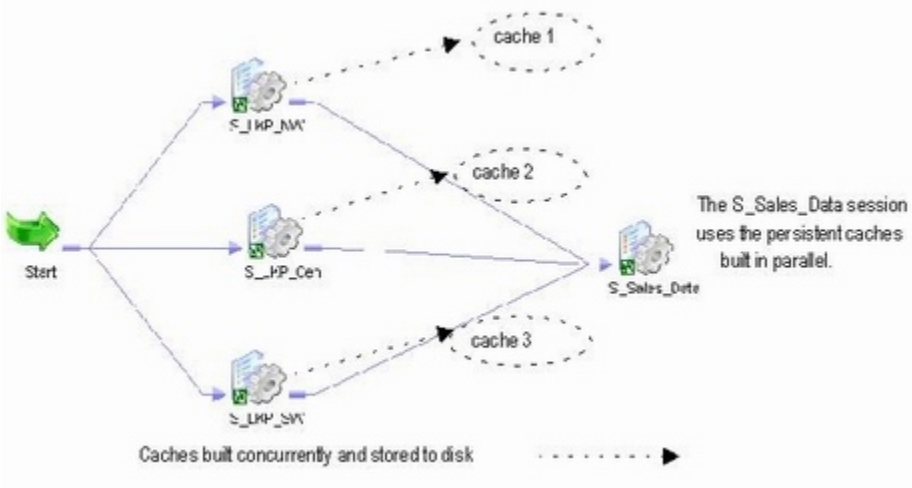
You create the session S_Sales_Data based on the mapping you created in Step 3. This session processes the data.

Step 5: Create a Workflow to Include All the Sessions

Create a workflow that uses all the sessions. The Integration Service runs the three cache-building sessions in parallel. Therefore, it creates the persistent caches concurrently and improves performance. For example, if each cache took

fifteen minutes to build, it would take forty-five minutes to build the caches sequentially, but only fifteen minutes to build the caches concurrently.

The session that processes the data runs after the sessions that build the caches. The Integration Service builds the memory cache for this session from the named persistent cache files. This improves performance because the Integration Service does not have to create the cache from the database.



Note: If you need to rebuild caches each time you run the session, you can specify the Recache from Lookup Source property when you create the mappings. If you do not need to rebuild the caches each time you run the session, you can use this workflow to build caches, and run a separate workflow to process the data thereafter.

Author

David Meister
Technical Support Engineer

David Meister is a four-year veteran of Informatica Global Customer Support. His current interests are Unstructured Data, Data Transformation, and B2B Data Exchange.